設計・製造におけるチップの脆弱性検知手法の研究開発
Research and Development of AI-Based Chip Vulnerability Inspection Methods in Design and Manufacturing（AVIM）

代表研究責任者　戸川 望（早稲田大学）

研究開発期間　　令和元年度

【Abstract】

This project, Research and Development of AI-Based Chip Vulnerability Inspection Methods in Design and Manufacturing, aims to develop a novel machine learning method which would find characterizing features of hardware (HW) trojans.　This in turn would allow for an AI-based HW trojan detection system to distinguish HW trojans from a normal (trojan-free) circuit.

The methods to be developed are twofold:

1. Detection from designed circuit
   To detect HW trojans using circuit information such as logical gates, flip-flops, and connections between input and output terminals.
2. Detection from manufactured circuit
   To detect HW trojans focusing on circuit behavior such as power consumption and processing time.

The developed methods/systems are to be evaluated under several practical scenarios, such as internal inspection before production, security certification for HW products, acceptance inspection of HW products and monitoring for running HW products.

1　研究開発体制

○　**代表研究責任者**　　戸川 望（早稲田大学）

○　**研究分担者**

【KDDI 総合研究所】　　清本 晋作（KDDI 総合研究所）

【ラック】　　船引 裕司（ラック）

○　**総合ビジネスプロデューサ**　冲中 秀夫（Techno-Visionary）

○　**ビジネスプロデューサ**　　橋本 和夫（早稲田大学）

杉山 敬三（KDDI 総合研究所）

三木 俊明（ラック）

○　**研究開発期間**　令和元年度

○　**研究開発予算**　　総額　198,623 千円

（内訳）

令和元年度
198,623 千円

2　研究開発課題の目的および意義

Society 5.0 は、サイバー空間とフィジカル空間が高度に融合したサイバーフィジカルシステムにより実現される。全ての「モノ」がネットワークに接続され、その電子機器の数は、2020 年には 400 億を超えると言われている。新しい価値やサービスが次々と創出され人々に豊かさをもたらす一方で、複雑化するサプライチェーン全体のセキュリティの確保は重要な課題となっている。

電子機器のハードウェア上に組み込まれた不正なチップは、製品出荷後に交換・修正することが難しく、その影響は極めて深刻になる可能性があることから、サプライチェーン上の脅威となっている。また、チップに仕込まれた不正な回路や部品を検出する技術は確立しておらず、産学官で連携して研究開発を加速し、社会実装を進めることが急務となっている。

以上の背景のもと、本研究開発ではハードウェアチップの設計・製造における脆弱性検知手法を確立するとともに、当該技術の社会実装を加速し、サプライチェーン全体のセキュリティ確保に資することを目的とする。

3　研究開発成果（アウトプット）

本報告では、全体 4 年間（令和元年度から令和 4 年度）の研究計画のうち、初年度・令和元年度の研究開発成果（アウトプット）を報告する。

課題 I として、まず論理ゲートやフリップフロップ、入出力端子等の接続情報によって与えられる回路情報（ネットリストと呼ばれる）をもとに、入出力線数や入出力端子・記憶素子からの論理段数等、不正回路を構成する信号線と不正でない回路を構成する信号線を有意に識別する特徴量を抽出する技術を研究開発する。そして、抽出された特徴量を機械学習することで、不正回路を構成する信号線と不正でない回路を構成する信号線を識別する技術を研究開発する（図 3-0-1）。

課題 I では、ア）不正回路を識別するための特徴量抽出技術に関する研究開発と、イ）AI/機械学習に基づく不正回路検知技術に関する研究開発を実施する。課題 I －ア）は、さらに①不正回路を識別するための特徴量抽出技術に関する要素技術開発と②設計・製造におけるチップの脆弱性検知手法に関する動向調査を実施する。

課題 I －ア）①では、Trust-HUB や ISCAS 回路等の（不正回路を含む）標準ベンチマークに対して、まず信号線自体の特徴量（自己特徴量）の抽出に取り組む。そして、抽出した特徴量のもとに、不正でない回路を構成する信号線を不正と判定する誤検知率が 5%以下という条件のもと、不正回路を構成する信号線を検知する確率を最大化する特徴量の組合せを抽出するものとする。特徴量を抽出する際、いかに「過学習」を回避するかが大きな問題となるが、本研究開発では、過学習することなく「最適な特徴量の抽出」を実現するために、ランダムフォレスト識別器による特徴量の「寄与度」やニューラルネットワークのドロップアウトの活用等を取り入れることで、これを実現する。

課題 I －ア）②では、チップの脆弱性検知手法技術の社会実装の加速とその業界・国際標準化等による

我が国の国際競争力強化を果たすため、技術に関する研究開発と並行して不正回路問題に対する諸外国の動向を調査する。

課題Ⅰ－イ）では、最新の不正回路の動向について調査を行い、不正回路の特徴、有効化方法の体系化を完了する。体系化の結果に基づき、代表的な不正回路を選定し、FPGA 開発ボードにおける不正回路のサンプルの実装および亜種生成ソフトウェアの開発を行う。不正回路のサンプル実装に関しては、FPGA 開発ボードや FPGA を搭載する電子機器等を対象とし、4 種類以上の不正回路のサンプルを実装する。亜種生成ソフトウェアは、入力された不正回路の回路情報に対し、論理的に等価であり、かつ異なる論理ゲートの構造を持つ不正回路の回路情報を自動生成する。

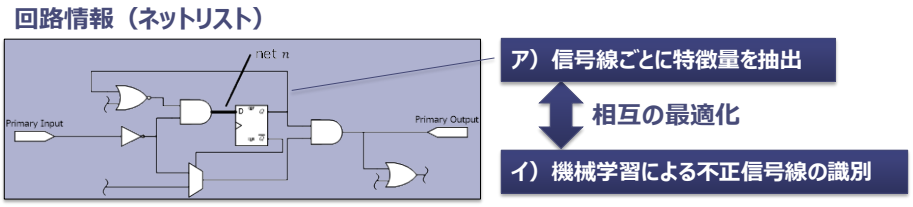


図 3-0-1：研究開発の概要（課題Ⅰ）

続いて、課題Ⅱとして、まず不正回路が組み込まれたチップにより構成される電子機器に対し、その電力波形の特定部分の電力量や継続時間、電力波形の時系列変化等を観測することで、不正回路により引き起こされる不正動作と、正常動作とを有意に識別する特徴量を抽出する技術を研究開発する。そして、抽出された特徴量を機械学習することで、電子機器の不正動作と正常動作とを識別する技術を研究開発する。

課題Ⅱでは、ア）外部情報を取得する電子機器の動作のモデル化技術と、イ）AI/機械学習に基づく不正動作検知技術に関する研究開発を実施する（図 3-0-2）。

課題Ⅱ－ア）では、まず組込みマイコン等のチップに不正回路が含まれていることを想定し、これらを用いて電子機器を構成した上で、その動作をモデル化する。ここで電子機器は必ず何らかの電力を消費することに注目する。そして、電子機器の直接的な動作が不明であったとしても、電子機器の動作はその電力波形を観測することで「通常通りの電力波形のパタンであるか」等を識別できる可能性が高い。しかも、電子機器の電力波形は、電子機器の電源供給端子を観測することで、他の外部情報に比較して、極めて容易に観測することができる。つまり、電子機器の外部情報として、電力波形を観測することは、

(1) すべての電子機器について、極めて容易に電力波形が取得できる、
(2) 電子機器の直接的な動作が不明であったとしても、通常の電力波形のパタンと異常な電力波形のパタンの識別可能、

といった大きな利点がある。そこで、特定の動作モデルを想定し、その外部情報として「電力波形」を取り上げ、電力波形を外部より観察することでこれを学習し、不正回路による不正動作と、通常動作を識別することを目指す。具体的に、いくつかの種類の組込みマイコンを対象とし、センサアプリケーション・通信アプリケーションを動作させる。アプリケーションプログラムに不正動作のシナリオを準備し、動作をモデル化する。電力波形を学習・解析することで不正動作と正常動作を識別することを目指す。

課題Ⅱ－イ）では、課題Ⅱ－アの研究においては専用の計測装置を用いる事が一般的であった。その導入に伴って数十万円～数百万円といった費用が必要となり、その計測装置自体が様々な応用用途での測定を可能とする多様な機能を持っているため、本研究のような電力データを測定するために複雑な機

能の計測装置を電力データの測定用に設定する作業があった。今日において、マイクロプロセッサの小型化ならびに A/D コンバーターの高精度化に伴い、電力データの測定に必要なハードウェアを構成できると考えた。小型化、軽量化、専用用途化する事によって専用に計測装置が不要であり可搬性にも優れた回路基板を用意する。ただし、この回路基板において電子機器の特徴量を捉える事が可能である必要がある。この点について事前調査の結果、電子機器に内蔵されるマイクロプロセッサの処理速度に対して周辺デバイスの応答について、マイクロプロセッサ上で動作する OS の最小処理時間が 1ms である事から、処理の特徴的波形が 1ms 単位（すなわち 1kHz 刻み）で発生する事が観測され、サンプリング精度 1kHz という目標を設定した。また、計測誤差については、電子機器の計測中の電流変動量を見た場合に±0.1mW 程度であれば十分に特徴量を捉えられるという事前研究のデータがあった。以上から±0.1mW の精度での電力の測定及び 1kHz 刻みでのサンプリング性能を目標とした高精度アナログ計測モジュールの検討を行い、計測モジュールを試作する。具体的なアプローチとしては、「現状のハードウェアの攻撃手法調査」および「PoC の実装となるハードウェアに関する最終数値目標（正常動作を不正動作と判定する誤検知率が 5%以下という条件のもと、不正動作を見逃す見逃し確率 10%以下を実現する）に対する実装の検討」を進める。また、単独の組込みマイコン等比較的容易な電子機器の動作のもと、見逃し確率を最小化するように、電力波形を用いて不正動作を検知する技術の基本検討を行う。

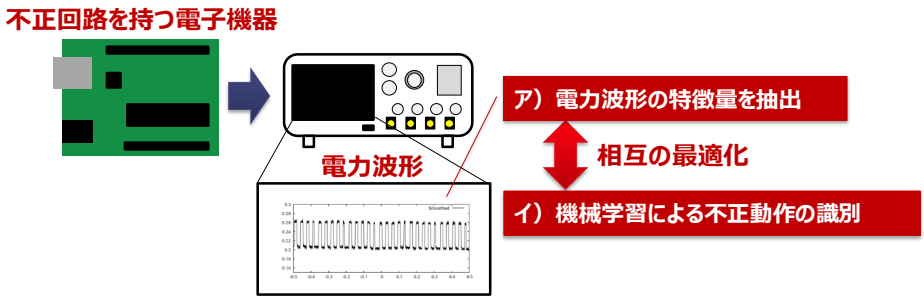


図 3-0-2：研究開発の概要（課題Ⅱ）

次項から本年度における各研究開発の進捗・成果を詳述する。本年度は以上の研究開発目標がすべて達成された（表 3-0-3）。

表 3-0-3　令和元年度における研究開発目標の達成状況

	令和元年度（1 年目）の目標	達成状況
課題 I	ア）不正回路を識別するための特徴量抽出技術 （1 年目）回路設計に標準的なベンチマーク回路等を用いて、不正回路の種類及びその機能を明確化し、不正回路と不正でない回路を識別するための特徴量を抽出する技術を開発する。	達成 (3.1 節・3.2 節)
	イ）AI/機械学習に基づく不正回路検知技術 （1 年目）ベンチマーク回路等を用いて、AI により回路情報から不正回路を検知する技術を開発する。	達成 (3.3 節)

課題Ⅱ	ア） 外部情報を取得する電子機器の動作のモデル化技術 （１年目）単独の組込みマイコン等、比較的簡易な電子機器の動作のもと、見逃し確率を最小化するような電子機器の外部情報の特徴量を抽出する技術を開発する。	達成 (3.4 節)
	イ） AI/機械学習に基づく不正動作検知技術 （１年目）単独の組込みマイコン等、比較的簡易な電子機器の動作のもと、見逃し確率を最小化するように、AI により外部情報から不正動作を検知する技術の基本検討を行う。	達成 (3.5 節)

３．１　課題Ⅰ－ア）① 不正回路を識別するための特徴量抽出技術に関する要素技術開発

ア）不正回路を識別するための特徴量抽出技術
（１年目）回路設計に標準的なベンチマーク回路等を用いて、不正回路の種類及びその機能を明確化し、不正回路と不正でない回路を識別するための特徴量を抽出する技術を開発する。

課題Ⅰ－ア）①では、上記目標を達成するため、さまざまな種類や機能の不正回路を含む標準ベンチマーク回路を用いて、不正回路と不正でない回路を識別するための特徴量抽出を達成した。特に回路情報に含まれる信号線自体の特徴量（自己特徴量）に注目し、不正回路と不正でない回路を識別するための有意な特徴量を見出した。標準ベンチマーク回路（Trust-HUB ならびに ISCAS ベンチマーク回路）を元に、不正でない回路を不正と判定する誤検知率が 5%以下という条件のもと、不正回路を見逃す見逃し確率 20%以下を実現する特徴量抽出技術を確立することを目標とした。

（１）信号線の自己特徴量の抽出

課題Ⅰでは、ハードウェア記述言語などによってゲートレベルネットリスト（プライマリ入出力ならびにゲートと信号線（ネットと呼ばれる）の接続によって与えられる）としてチップの設計情報が与えられた際、ネットリスト中の各ネットが、ハードウェアトロイを構成するネットであるか（トロイネットと呼ぶ）、あるいは通常回路を構成するネットであるか（通常ネットと呼ぶ）、に分類することを目指す。この際、各ネットが持つどのような特徴に基づき、ネットを分類するかが大きなポイントとなる。そこで、ネットの分類に必要な特徴を列挙することにする。ここでは特にネットそのものと、ネットの遷移的ファンインとなる特徴と遷移的ファンアウトとなる特徴に着目する（自己特徴量と呼ぶ）。

Trust-HUB で公開されているゲートレベルのネットリストを参照し、表 3-1-1 をもとに、ハードウェアトロイと関係する可能性が高いネットの特徴量として具体的に以下のものを取り上げた。ここで参照した Trust-HUB ベンチマーク回路は、

(1) 外部信号線を不正に改竄するもの
(2) チップ内部の情報を外部漏洩するもの
(3) 故意に回路性能を劣化させるもの

の 3 種類を選出している。

表 3-1-1 抽出した信号線の特徴量の候補（パラメータにより 51 種類）

特徴量	意味
fan_in_x	ネットｎの入力側x段手前に接続される論理ゲートの数
in_flipflop_x	ネットｎの入力側x段手前に接続されるフリップフロップの数
out_flipflop_x	ネットｎの出力側x段手前に接続されるフリップフロップの数
in_multiplexer_x	ネットｎの入力側x段手前に接続されるマルチプレクサの数
out_multiplexer_x	ネットｎの出力側x段手前に接続されるマルチプレクサの数
in_loop_x	ネットｎの入力側x段それぞれでループを構成する数
out_loop_x	ネットｎの出力側x段それぞれでループを構成する数
in_const_x	ネットｎの入力側x段手前に接続される定数の数
out_const_x	ネットｎの出力側x段手前に接続される定数の数
in_nearest_pin	ネットｎから最も近いプライマリ入力の段数
out_nearest_pout	ネットｎから最も近いプライマリ出力の段数
{in, out}_nearset_flipflop	ネットｎから入力/出力側で最も近いフリップフロップの段数
{in, out}_nearest_multiplexer	ネットｎから入力/出力側で最も近いマルチプレクサの段数

（2）信号線の自己特徴量の最適化

機械学習では、多くの特徴量を用いることが必ずしも識別率が高い識別結果を導出するとは限らない（一般的に「過学習」と呼ばれる)。そこで、上記（1）で信号線の自己特徴量に対して、ハードウェアトロイを構成するネットと通常回路を構成するネットの分類に、実際にどの特徴量が大きく寄与するかを評価した。

Random Forest の重要度を用いた自己特徴量の最適化

Random Forest とは機械学習アルゴリズムの一つであり、特徴としてクラス識別に使用した特徴量の重要度を算出することが可能となる。この重要度を利用することで、ハードウェアトロイ識別に有効な特徴量を抽出して最適な数に絞り込むことを考える。特徴量を選択する流れは大きく Step 1 と Step 2 に分かれる。Step 1 でネットリストから特徴量を算出し、Step 2 で実際に特徴量の絞り込みを行った。

Step 1：ネットリストから特徴量算出

Step 1 ではネットリストに対してそれぞれのネットの特徴量を算出する。ここで算出する特徴量はハードウェアトロイ識別に有用であるか否かに関係なく、表 3-1-1 にもとづいたネットの特徴を表す特徴量であり、回路を構成するネットリスト、すなわちネットの集合に対して、集合に含まれる各ネットnに対して、表 3-1-1 に基づき複数個の特徴量が得られる。各ネットnに対する特徴量をベクトル$\overrightarrow{v_n}$とする。

Step 2：Random Forest による特徴量の最適化

Step 2 では Random Forest を用いて学習することで、特徴量を最適化する。Random Forest では、すべての特徴量に対して重要度が同時に算出されることに注目する。Trust-HUB から 15 個のベンチマーク回路を抽出し、表 3-1-1 にある 51 個のすべての特徴量を用いて、一個抜き交差検証（評価対象の 1 つの回路以外を学習し、評価対象の回路を未知回路として、回路中の信号線のトロイ／非トロイを識別する）による識別を行った。この際、重要度が高いものからまず 25 個を取り出した。同様に 25 個の特

徴量を用い、一個抜き交差検証による識別を行った。なお、このとき、25 個の特徴量を用いた識別結果は、51 個の特徴量を用いた識別結果よりも識別率は高かった。なお Step2 では識別率として F-measure を用いている。

さらにその中から重要度が高いものを 12 個取り出す、等の工程を繰り返すことで、最初の 51 個の特徴量を用いたときと識別率を同等以上に保持したまま最終的に表 3-1-2 に示す 11 種類の特徴量を抽出した。以降、この 11 種類の特徴量を用いることにする。

表 3-1-2 最適化された特徴量（11 種類）

特徴量	重要度	特徴量	重要度
fan_in_5	0.102	out_loop_5	0.104
in_flipflop_4	0.040	in_nearest_pin	0.108
in_flipflop_5	0.065	out_nearest_pout	0.207
out_flipflop_3	0.125	out_nearest_flipflop	0.048
in_loop_4	0.062	out_nearest_multiplexer	0.069
in_loop_5	0.070		

ニューラルネットワークのドロップアウトを用いた自己特徴量の最適化

さらに表 3-1-1 の 51 種類の特徴量をニューラルネットワーク（中間層を 1 層～3 層までさまざまな数で変化させた）に入力し、ドロップアウトを適用した結果、上記表 3-1-2 の 11 種類の特徴量による識別結果が最良となった。

つまり、表 3-1-2 の 11 種類の特徴量が現段階では最適な特徴量の集合と結論づけられる。

（3）機械学習による識別結果

上記によって最適化された 11 個の特徴量を利用して、Trust-HUB 中の回路ならびにハードウェアトロイを含む ISCAS89 ベンチマーク回路（ISCAS89 ベンチマーク回路に Trust-HUB によるハードウェアトロイを組み込んだもの）に加え、Trust-HUB 中の実際の IP コア（Intellectual Property コア、10 万個を超える信号線を持つ）を対象に、課題 I－イ）と共同で Random Forest ならびに多層ニューラルネットワークによるハードウェアトロイの識別を行った。図 3-1-1 に識別評価の方法を示す。ここでも一個抜き交差検証を行った。

また Random Forest について、弱学習器の数を 200、不純度の計算を交差エントロピー、木の深さの最大を 13、最小分割数を 2 とした。多層ニューラルネットワークについて、入力層 11、中間層 200、100、50、出力層 2 とした。これらのハイパーパラメータは、予備実験を通し、最も識別率が高いものを選択している。

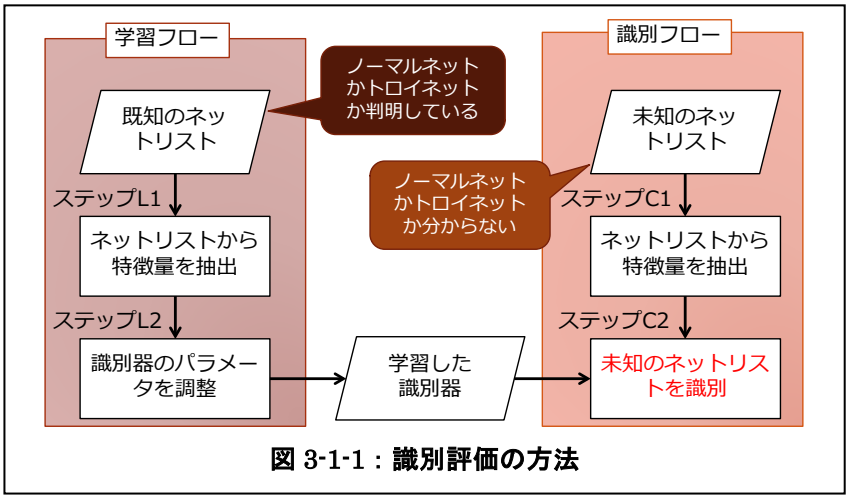


図 3-1-1：識別評価の方法

Random Forest による識別結果を表 3-1-3 に、ニューラルネットワークによる識別結果を表 3-1-4 に示す。表中、真のノーマルネットのうち正しくノーマルネットと識別したものの数を TN、誤ってトロイ

ネットと識別したものの数を FP とする。真のトロイネットのうち正しくトロイネットと識別したものの数を TP、誤ってノーマルネットと識別したものの数を FN とする。TPR はトロイネットを正しくトロイネットとして識別した割合であり、TP/(TP＋FN)で表される。TNR はノーマルネットを正しくノーマルネットとして識別した割合であり、TN/(TN ＋ FP)で表される。Accuracy は(TP+TN)/(TN+FP+FN+TP)で表される。

なお「不正でない回路を不正と判定する誤検知率」は(1-TNR)で表すことができ、「不正回路を見逃す見逃し確率」は(1-TPR)で表すことができる。研究開発運営委員会による議論等を通して、本研究開発における令和元年度の目標は、TNR が 95%以上という条件のもと、TPR を 80%以上にすることにした。

表 3-1-3 と表 3-1-4 から、表 3-1-2 の 11 種類の特徴量を用いることで、ISCAS89 や Trust-HUB 中の回路を表すネットリストについて、特に多層ニューラルネットワークによるネット識別によって、平均 TPR が 84.6%、平均 TNR が 95.1%であることから、多様なベンチマーク回路において平均的に、不正でない回路を不正と判定する誤検知率を 5%以下にした上で、不正回路を見逃す見逃し確率を 15%強程度にすることを達成した。すなわち、課題Ⅰ－ア）①および課題Ⅰ－イ）において、令和元年度の目標を達成した。

なお、Random Forest による識別結果は、FP がほぼ 0 になっていることが分かる。これは、回路中にハードウェアトロイを構成する信号線が含まれていれば、ほぼ誤りなくトロイ信号線だけを発見できることを表している。ハードウェアトロイ信号線をすべて検出することが目的でなく、単に回路にハードウェアトロイが含まれているかどうかを検知するには、Random Forest による識別が有利であることを意味している。開発技術を用いた実用化シナリオを考える上で、識別器の特性による振る舞いが大きな意味を持つと考えられる。

表 3-1-3 Random Forest を用いた識別結果

分類する回路	TN	FP	FN	TP	TPR	TNR	Accuracy
RS232-T1000	309	0	0	10	100.0%	100.0%	100.0%
RS232-T1100	309	0	0	11	100.0%	100.0%	100.0%
RS232-T1200	310	0	3	10	76.9%	100.0%	99.1%
RS232-T1300	309	0	2	5	71.4%	100.0%	99.4%
RS232-T1400	306	0	0	12	100.0%	100.0%	100.0%
RS232-T1500	311	0	0	11	100.0%	100.0%	100.0%
RS232-T1600	311	0	6	4	40.0%	100.0%	98.1%
s15850-T100	2420	0	24	2	7.7%	100.0%	99.0%
s35932-T100	6408	0	14	0	0.0%	100.0%	99.8%
s35932-T200	6405	0	11	1	8.3%	100.0%	99.8%
s35932-T300	6405	0	24	13	35.1%	100.0%	99.6%
s38417-T100	5799	0	11	0	0.0%	100.0%	99.8%
s38417-T200	5802	0	11	0	0.0%	100.0%	99.8%
s38417-T300	5801	0	13	31	70.5%	100.0%	99.8%
s38584-T100	7343	0	19	0	0.0%	100.0%	99.7%
s38584-T200	7362	3	85	12	12.4%	100.0%	98.8%
s38584-T300	7615	0	870	3	0.3%	100.0%	89.8%
EthernetMAC10GE-T700	102969	0	3	9	75.0%	100.0%	100.0%
EthernetMAC10GE-T710	102969	0	2	10	83.3%	100.0%	100.0%
EthernetMAC10GE-T720	102969	0	2	10	83.3%	100.0%	100.0%
EthernetMAC10GE-T730	102969	0	11	1	8.3%	100.0%	100.0%
B19-T100	70968	1	0	96	100.0%	100.0%	100.0%
B19-T200	70968	1	0	96	100.0%	100.0%	100.0%
wb_conmax-T100	21162	3	11	0	0.0%	100.0%	99.9%
Average	-	-	-	-	48.9%	100.0%	99.3%

表 3-1-4 ニューラルネットワークを用いた識別結果

分類する回路	TN	FP	FN	TP	TPR	TNR	Accuracy
RS232-T1000	299	10	0	10	100.0%	96.8%	96.9%
RS232-T1100	298	11	0	11	100.0%	96.4%	96.6%
RS232-T1200	301	9	0	13	100.0%	97.1%	97.2%
RS232-T1300	290	19	0	7	100.0%	93.9%	94.0%
RS232-T1400	300	6	0	12	100.0%	98.0%	98.1%
RS232-T1500	295	16	0	11	100.0%	94.9%	95.0%
RS232-T1600	301	10	0	10	100.0%	96.8%	96.9%
s15850-T100	2240	180	5	21	80.8%	92.6%	92.4%
s35932-T100	6308	100	6	8	57.1%	98.4%	98.3%
s35932-T200	6050	355	7	5	41.7%	94.5%	94.4%
s35932-T300	6367	38	0	37	100.0%	99.4%	99.4%
s38417-T100	5712	87	2	9	81.8%	98.5%	98.5%
s38417-T200	5491	311	1	10	90.9%	94.6%	94.6%
s38417-T300	5482	319	0	44	100.0%	94.5%	94.5%
s38584-T100	7269	74	16	3	15.8%	99.0%	98.8%
s38584-T200	6253	850	39	58	59.8%	88.0%	87.7%
s38584-T300	7220	395	95	579	85.9%	94.8%	94.1%
EthernetMAC10GE-T700	102366	603	4	8	66.7%	99.4%	99.4%
EthernetMAC10GE-T710	102524	445	1	11	91.7%	99.6%	99.6%
EthernetMAC10GE-T720	102463	506	0	12	100.0%	99.5%	99.5%
EthernetMAC10GE-T730	101212	1757	5	7	58.3%	98.3%	98.3%
B19-T100	63275	7374	0	96	100.0%	89.6%	89.6%
B19-T200	63700	6949	0	96	100.0%	90.2%	90.2%
wb_conmax-T100	17188	4998	0	11	100.0%	77.5%	77.5%
Average	-	-	-	-	84.6%	95.1%	95.1%

３．２　課題Ｉ－ア）② 設計・製造におけるチップの脆弱性検知手法に関する動向調査

一般に、ハードウェアのサプライチェーンセキュリティにおける問題として、模造集積回路(Counterfeit integrated circuits (ICs))が指摘されている。模造集積回路は、ハードウェアトロイ、リサイクルチップ、仕様外／欠陥チップ、クローンチップに分類される。ハードウェアトロイは、チップが改ざんされたものであり、後述する通り、その挙動や特徴はさまざまなものがある。リサイクルチップは廃棄された電子部品を再生したものであり、従前の使用により既に機能しなくなっている、または、性能が衰えているなどの脆弱性を持つ。仕様外／欠陥チップは、必ずしも悪意のある脆弱性ではなく、単なるバグ等が混入したチップである。クローンチップは典型的な「海賊版」であり、正規品がコピーされたものである。

これら広く模造集積回路を対象に、サプライチェーンをセキュア化する仕組み（標準や認証）がある。もっとも、本研究開発における動向調査の結果、それらはすでに参考資料（いわゆるゴールデンチップ）と比較するものや、光学的検査を利用するものなどであり、ハードウェアトロイのような改ざんチップの検知に十分対応できるものではないことが分かった（Common Criteria (CC) (ISO/IEC 15408) や、Society of Automotive Engineers (SAE) International・AS6171A、Components Technology Institute (CTI)・CTI CCAP-101、The Independent Distributors of Electronics Association (IDEA)・IDEASTD-1010)。

そのため、特にハードウェアトロイに着目した場合に、チップのサプライチェーン上での安全性を確保するため、どういった脅威を対象とするかが問題となる。そこで、本年度、本研究開発の実用化・社会実装の方向性を検討することとし、具体的には、①課題Ⅰ－イ）における Trust-HUB を活用した不正回路の機能に基づき、②ハードウェアトロイがサプライチェーンセキュリティに与える影響を分析し、国内外の法令・標準を参考に、本研究開発において対象とするチップのサプライチェーン上の脅威を検討した。

① チップサプライチェーンセキュリティの観点によるハードウェアトロイの大分類

ハードウェアトロイの機能は大きく、(1)情報の機密性に関するものと(2)機能の変更・妨害を行うものに分類できる。(1)情報の機密性に関するハードウェアトロイは、回路内部にアクセス（読み／書き）することで、チップ内部に保持された機密性を損なうもので、例えば、回路内部に秘密情報を持っている場合、回路内部にアクセスし、これらを漏洩するものである。(2)機能の変更・妨害を行うハードウェアトロイは、必ずしも情報の機密性を損なうことはないが、ハードウェアトロイによって回路機能が改ざんされ、チップの安全性を損なう可能性がある。たとえば、特定の条件のもとハードウェアトロイのペイロード回路が機能し、その結果、チップの特定の機能が制限される／無効化される、あるいは仕様と異なる動作を行うものである。

② チップサプライチェーンセキュリティの観点によるハードウェアトロイの脅威

①で表されるように、ハードウェアトロイの機能には異なった脅威があるため、チップのサプライチェーンに与える影響も異なってくる。

まず、(1)情報の機密性に対する脅威について検討する。サプライチェーン上流でチップを利用するメーカは製品やサービスの提供者に無過失の責任を負うため（製造物責任法 3 条や個人情報保護法 20 条など)、サプライチェーンの法的なリスクになる。実際、米国の政府機関向けの暗号化モジュールに関するセキュリティ要件 NIST・FIPS 140-2 は、レベル 4 では遮蔽措置によるサイドチャネル攻撃からの防御措置を講じることを求めており、ハードウェアにおける情報の機密性に対処することが求められている。そこで、本研究開発においても、情報の機密性に対する脅威を対象にする必要がある。

次に、(2)機能の変更・妨害を行うものに対する脅威について検討する。サプライチェーン上流のメーカは法制度上、製品やサービスの提供者に無過失の義務が課されるため（製造物責任法 3 条)、ここでもサプライチェーンのリスクになる。実際、NIST・FIPS140-2 ではレベル 2・3 で改ざん防止の物理的セキュリティが要求されており、ハードウェアにおける情報の改ざんに対処することが求められている。そこで、本研究開発においても、機能の変更・妨害等による情報の改ざんひいては安全性に対する脅威を対象にする必要がある。

以上の通り、サプライチェーン上のハードウェアトロイの脅威が明らかになったことから、これらの脅威に基づく研究開発を実施するものとする。例えば、(1)情報の機密性に関するハードウェアトロイでは、回路内部にアクセス可能とする回路の特徴の一例として、レジスタやメモリなどの記憶素子から比較的短い段数で外部入出力に接続される。そこで、実用化・社会実装に向け、こうした回路構造の特徴を機械学習することで、機密性を損なう回路ならびにその派生回路を検知する技術を確立する。次に、(2)機能の変更・妨害を行うハードウェアトロイでは、回路機能の改ざんを行う回路の特徴の一例として、特定のトリガ条件を構成するために、チップ中の多数の箇所から信号線を集中させ、条件を構成する回路構造がある。そこで、実用化・社会実装に向け、こうした回路構造の特徴を機械学習することで、機能の変更・妨害を行う回路ならびにその派生回路を検知する技術を確立する。

なお、上記の情報の機密性に関するハードウェアトロイや機能の変更・妨害を行うハードウェアトロイは攻撃者が意図的に挿入したものか、あるいは仕様外／欠陥チップのように何らかの誤りで挿入されたものかを識別することは原理上不可能である。加えて、前述のとおり、法制度上、サプライチェーン上流のメーカの法的責任は無過失責任であり、攻撃者の意図の有無に応じて免責されない。そこで、設計者の故意あるいは単純な誤りであっても、チップに埋め込まれた機密性／機能の変更・妨害に関わるハードウェアトロイを発見する技術開発が必要となる。

上記のように、実用化・社会実装の観点から、チップサプライチェーン上で実際に脅威となるハードウェアトロイを想定した上で、これらに基づくハードウェアトロイならびにその派生回路を設計し、検知する技術の研究開発が必要となることを明らかにした。

３．３　課題Ⅰ－イ）AI/機械学習に基づく不正回路検知技術

イ）AI/機械学習に基づく不正回路検知技術
（１年目）ベンチマーク回路等を用いて、AI により回路情報から不正回路を検知する技術を開発する。

3.1 節に示した通り、課題Ⅰ－ア）①と連携して、AI の一種である Random Forest 及び多層ニューラルネットワークを用いて回路情報から不正回路を検知する技術を開発した。特に多層ニューラルネットワークによるネット識別によって、平均 TPR が 84.6%、平均 TNR が 95.1%であることから、多様なベンチマーク回路において平均的に、不正でない回路を不正と判定する誤検知率を 5%以下にした上で、不正回路を見逃す見逃し確率を 15%強程度にすることを達成した。すなわち、課題Ⅰ－イ）において、令和元年度の目標を達成した。

不正回路検知技術のさらなる高度化に向けたては、ベンチマーク回路だけではなく、実回路による評価が不可欠である。しかしながら、具体的な不正シナリオに基づく不正回路のサンプルは、マルウェア等のように入手することは困難である。そこで、本年度は上記の活動に加え、AI による不正回路の検知技術の高度化に向けた調査及び不正回路サンプルの実装、亜種生成ツールの開発も実施した。具体的には、最新の不正回路の動向について調査を行い、不正回路の特徴、有効化方法の体系化を完了した。体系化の結果に基づき不正回路の分類を網羅するように、IoT 開発ボード、FPGA 開発ボードおよび FPGA 搭載ネットワークボードを対象とし、計 12 種類の不正回路のサンプルを実装した。また、入力された不正回路の回路情報（Verilog-HDL で記述されたゲートレベルネットリスト）に対し、論理的に等価であり、かつ異なる論理ゲートの構造を持つ不正回路の回路情報を自動生成する亜種生成ツールも開発した。

（１）不正回路の分類

Trust-HUB のウェブサイトから 88 個の不正回路を入手し、入手した回路を有効化方法（トリガ回路）と不正回路（ペイロード回路の動作）の動作により分類した。

トリガ回路による分類としては、まず、トリガ回路が無いものと有るものに分類できる。トリガ回路が無いものについては、不正回路は常に有効化されている。トリガ回路が有るものについては、特定の条件が満たされることで、不正回路の機能（ペイロード回路）がはじめて有効化されるものである。トリガ回路が有る不正回路については、内部状態に基づくトリガ回路を持つものと、外部入力に基づくものに大別できる。内部状態に基づくトリガ回路としては、時間経過により起動するものと物理的条件により起動す

るものがある。外部入力に基づくトリガ回路としては、利用者の入力によるものと他の構成要素の出力によるものがある。

ペイロード回路による分類としては、回路から情報を漏洩せるもの、回路の機能を変更するもの、回路の機能を妨害するものに大別できる。情報を漏えいする回路の例としては、秘匿すべき鍵情報を秘密裏に出力するものがある。機能を変更する回路の例としては、出力値を改ざんするものがある。機能を妨害する回路は、機能の無効化を目的とするものと性能の低下を妨害するものに分類できる。前者の例としては、回路による処理を回避し、入力をそのまま出力するものがある。後者の例としては、冗長な回路より、回路全体の消費電力を増加させるものがある。

分類した結果を図に示す。

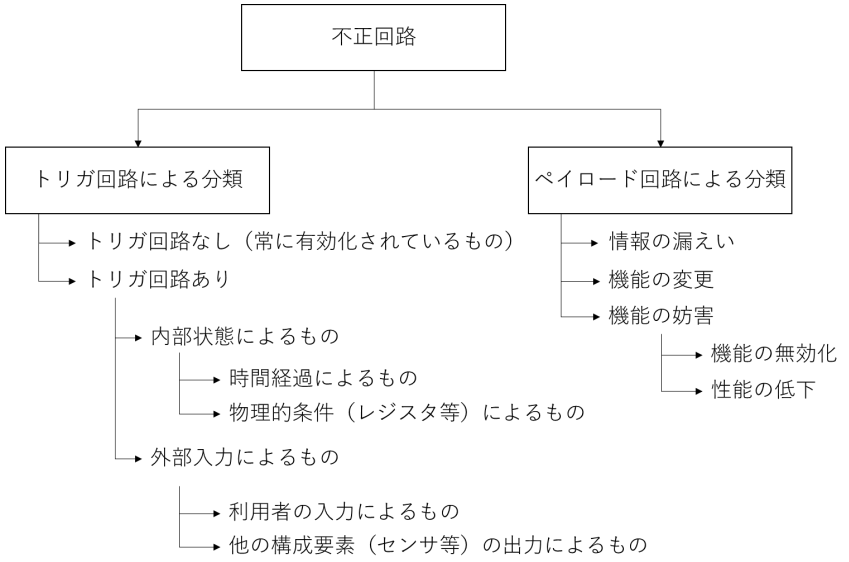


図 3-3-1　不正回路の分類

（2）不正回路のサンプルの実装

IoT 開発ボード、FPGA 開発ボード及び FPGA 搭載ネットワークボードを対象に、合計 12 個の新規の不正回路のサンプルを実装した。

1. IoT 開発ボード向けの不正回路サンプル

Lattice 社の iCEblink40-LP1K Evaluation キットを実装対象とした。ボードに接続した照度センサから照度データ（0～255）を定期的に取得し、USB ケーブルで接続した PC 上のソフトウェアで表示する回路を通常回路として実装した。不正回路としては、下表に示す 6 種類を実装した。

表 3-3-1　実装した IoT 開発ボード向け不正回路サンプル

項番	トリガ回路	ペイロード回路
1	時間経過による起動	センサの出力を無効化 （回路機能の妨害を目的）
2	時間経過による起動	センサの出力を最大値（255）に改ざん
3	複数ボタンの同時押下による起動	センサの出力を無効化
4	複数ボタンの同時押下による起動	センサの出力を最大値（255）に改ざん

5	指定順序でのボタンの押下による起動	センサの出力を無効化
6	指定順序でのボタンの押下による起動	センサの出力を最大値（255）に改ざん

2. FPGA 開発ボード向けのサンプル

Xilinx 社の Zynq ZC702 ボードを実装対象とした。利用者が指定した平文（16 バイト）と鍵（16 バイト）に基づき、128 ビット AES による暗号化処理を実行し、暗号文（16 バイト）を出力する通常回路を実装した。不正回路としては、下表に示す 3 種類を実装した。

表 3-3-2　実装した FPGA 開発ボード向け不正回路サンプル

項番	トリガ回路	ペイロード回路
1	特定のパタンを持つ平文により起動	出力（暗号文）を改ざん（回路機能の妨害を目的）
2	特定のパタンを持つ平文により起動	出力を改ざん（差分故障解析による鍵の漏えいを目的）
3	特定のパタンを持つ平文が、規定回数連続して入力される場合に起動	出力に鍵を埋め込む

3. FPGA 搭載ネットワークボード向けのサンプル

Digilent 社の NetFPGA 1G-CML ボードを実装対象とした。このネットワークボードは、FPGA および 4 つのイーサネットポートを有しており、ネットワークインタフェースカード、スイッチおよびルータの機能を実装できる。スイッチングハブの機能を提供する回路を通常回路として実装した。不正回路としては、下表に示す 3 種類を実装した。

表 3-3-3　実装したネットワークボード向け不正回路サンプル

項番	トリガ回路	ペイロード回路
1	特定のパタンを持つイーサネットフレームの受信により起動	半永久的に受信したイーサネットフレームを破棄（恒久的な回路機能の妨害を目的）
2	特定のパタンを持つイーサネットフレームの受信により起動	受信したイーサネットフレームを、指定された個数分破棄（一時的な回路機能の妨害を目的）
3	特定のパタンを持つイーサネットフレームの受信により起動	イーサネットフレームを宛先ポート以外のポートにミラーリング（情報の漏洩を目的）

（3）亜種生成ツールの開発

不正回路の亜種を自動生成する亜種生成ツールを開発した。不正回路亜種生成ツールは、不正回路の回路情報（Verilog-HDL で記述されたゲートレベルネットリスト）を入力とし、その不正回路の亜種の回路情報（Verilog-HDL で記述されたゲートレベルネットリスト）を出力する。亜種の生成方法としては、論理回路の等価変換により行う。具体的には、1) A NAND B = NOT(A) OR NOT(B)や 2) NOT(NOT(A)) = A、などの変換規約を用いる。前者の変換規約により、1 つの 2 入力の NAND ゲートを、1 つの 2 入力 OR ゲートと 2 つの NOT ゲートから構成される等価な回路に変換できる。また、後者の規約により、任意の

信号線に対し、2 つの NOT ゲートを直列に挿入する等価変換が可能である。今回の亜種生成ツールにおいては、OR ゲート、AND ゲート、NOR ゲート、NAND ゲート、XOR ゲートおよび XNOR ゲートの 6 種類の論理ゲートに対し、それぞれ 15 種類の変換規則を定義している。信号線に対する 2 つの NOT ゲートの挿入による変換規則と合わせて、合計 91 個の変換規則を利用できる。

下図に、亜種生成ツールのスクリーンショットを示す。左側が利用者が入力した不正回路のゲートレベルネットリストであり、右側がツールにより自動生成された亜種のゲートレベルネットリストである。

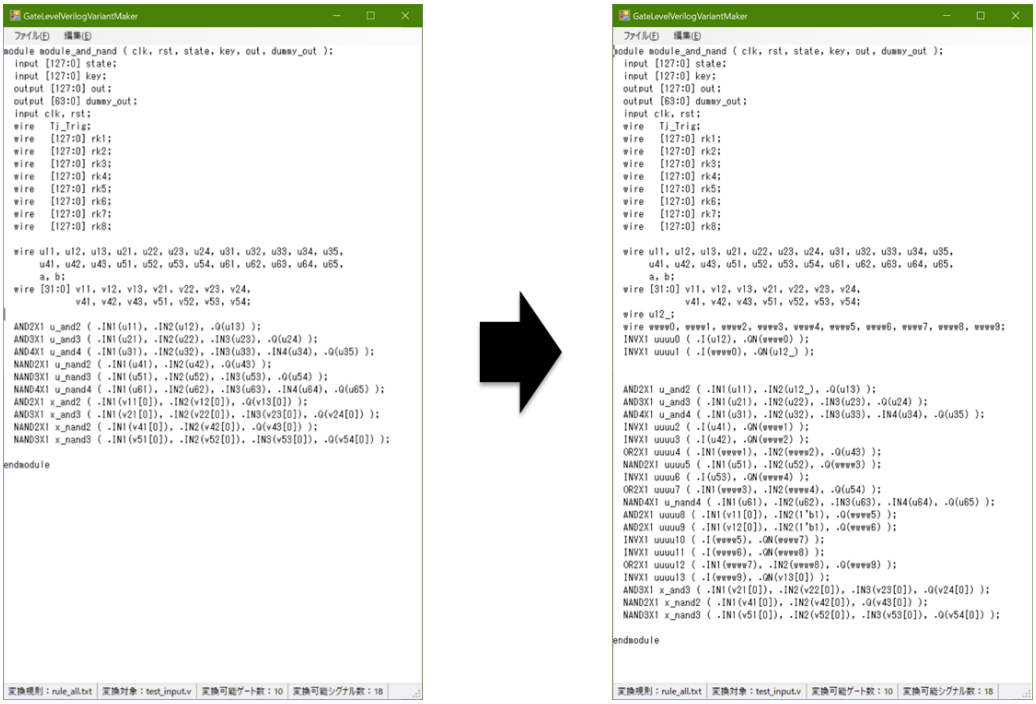


図 3-3-2　亜種生成ツールの画面例

３．４　課題Ⅱ－ア）外部情報を取得する電子機器の動作のモデル化技術

ア）　外部情報を取得する電子機器の動作のモデル化技術
（１年目）単独の組込みマイコン等、比較的簡易な電子機器の動作のもと、見逃し確率を最小化するような電子機器の外部情報の特徴量を抽出する技術を開発する。

課題Ⅱ－ア）では、組込みマイコン等に焦点を当て、組込みマイコン等のチップに不正回路が含まれていることを想定し、これらを用いて電子機器を構成した上で、その動作をモデル化した。そして動作モデルを用いて不正動作の検知を目指した。

（１）組込みマイコンの動作モデル化

組込みマイコンに挿入された悪意のある機能は、情報漏洩や機能の無効化を引き起こす。悪意のある機能は一般に、1)常に動作し続けるものと、2)特定の条件を満たした場合にだけ動作するものに分けられる。1)常に動作し続けるものは、マイクロコントローラが動作している間は常に悪意のある機能が動作する。これにより消費電力を増加させ、搭載機器の耐久性や信頼性を低下させる。2)特定の条件を満たした場合にだけ動作するものは、動作時刻やセンサ等の入力値などが特定の条件を満たした場合にだけ動作する。これにより、出荷時のテスト段階では悪意のある機能が隠蔽され、通常利用している間の任意のタイミングでだけ悪意のある機能が動作し、その時点の内部情報流出などを引き起こす。本研究開発では、令和元年度において、第一段階として 2)特定の条件を満たした場合にだけ動作するような悪意のある機能を対象とした。

特定の条件を満たした場合にだけ動作するような悪意のある機能の場合、短期間の動作テストでは悪意のある機能を検知することは難しい。そこで、工場内で実際の製品利用を想定した長期間の動作テストにおいて適用されることを想定する。

センサロガーなどの IoT 機器の多くは組込みマイコンを搭載する。これらの機器に搭載される組込みマイコンでは、電力の消費を抑えるため一般に通常モードとスリープモードの２つの動作モードを持つ。そこでこれらの動作をモデル化することを考える。

通常モード: 通常モードでは、組込みマイコンは通常の処理を実行する。例えばセンサロガーの場合、組込みマイコンはセンサから情報を取得し、その情報を内部メモリに記憶または外部に接続されたホストコンピュータへセンサ情報を送信する。通常処理では、マイクロコントローラは数十から数百 mW 程度の電力を消費する。

スリープモード: スリープモードでは、組込みマイコンの主要な機能を停止する。次回スリープモードから復帰するために最低限必要な機能以外の機能を無効化することで、電力消費を抑える。スリープモード中の組込みマイコンの消費電力は、通常モード中の消費電力と比較して非常に小さい。

図 3-4-1 に、想定する組込みマイコンの動作モードを示す。スリープモードでは、組込みマイコンでは主要な機能が停止し、タイマによるトリガを待つ。トリガがかかると動作モードに移行する。動作モードでは、通常の機能が動作する。ここで、悪意のある動作が組込みマイコンに挿入された場合、通常モードにおいて、ごく稀に発生するトリガ条件にもとづき悪意のある機能が動作すると仮定する。この場合、通常モード時の消費電力において、悪意のある機能が動作していない場合に比べ、通常モードの継続時間やその間の消費エネルギーに変化が生じる。この変化に着目し、悪意のある機能の発現を検知することを考える。

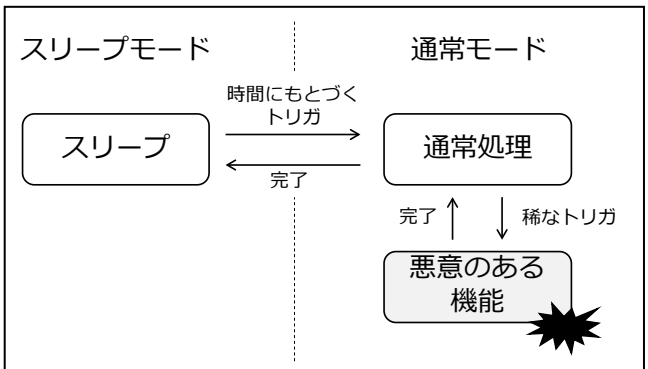


図 3-4-1：通常モードとスリープモードを持つ動作モデル

（2）消費電力を利用した異常動作の検知手法の確立

消費電力解析にもとづく、組込みマイコンに挿入された悪意のある機能の発現検知手法を開発した。以下の 5 つの処理から構成される（図 3-4-2 参照）。

(1) 電力測定: 測定機器を用いて対象となる電子機器の消費電力を計測する。

(2) 波形整形: 測定時のノイズを軽減するため、消費電力の測定波形を平滑化する。

(3) 通常/スリープモードの識別: 教師なし学習を用いて消費電力波形を通常モードの区間とスリープモードの区間に識別する。

(4) 特徴量抽出: 通常モードと識別された区間に対し、その継続時間とその間の消費エネルギーを算出し、特徴量として抽出する。

(5) 外れ値検知: 抽出された特徴量に対し外れ値検知アルゴリズムを適用し、異常な区間を検出する。検出された区間では、悪意のある機能が発現したと識別される。外れ値検知アルゴリズムでは局所外れ値因子（LOF）法を適用する。また特徴量として、上記(4)で抽出された通常モードのそれぞれの区間における(a)継続時間と(b)消費エネルギーの 2 値を用いることにした。

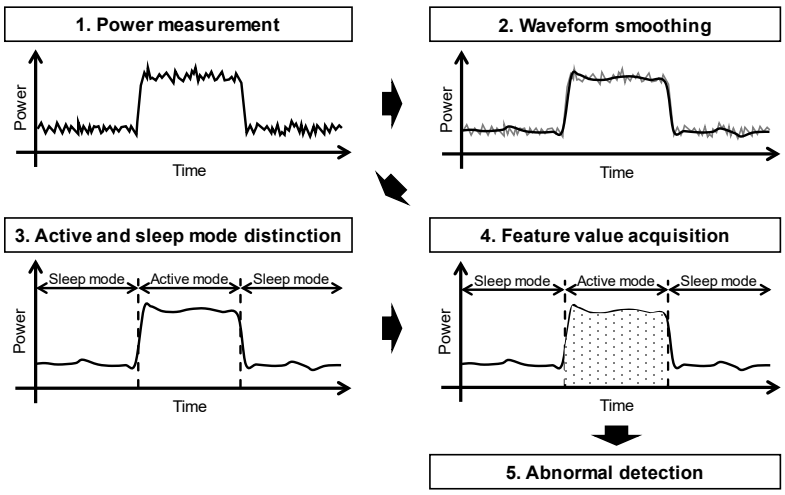


図 3-4-2：動作モデルに基づく異常動作の検知フロー

（3）Arduino UNO に関する異常動作検知

（2）の検知フローを用いて、8 ビットマイクロコントローラ Microchip ATmega328P を搭載した Arduino UNO ボードを用いて異常動作を検知した。アプリケーションとして暗号化機能をもったセンサロガーを想定した。図 3-4-3 に、アプリケーションの概要を示す。通常モードにおける主な機能として、1) AD 変換、2) AES 暗号化、3) シリアル出力がある。これらの処理を終えると、マイクロコントローラはスリープモードに移行する。実装したプリケーションでは、32ms ごとに通常モードに復帰する。

以上のアプリケーションに対し、悪意のある機能を挿入した。悪意のある機能では、5 回に 1 回 AES 暗号化処理を無効化し、AD 変換で得られた結果をそのままシリアル出力する。

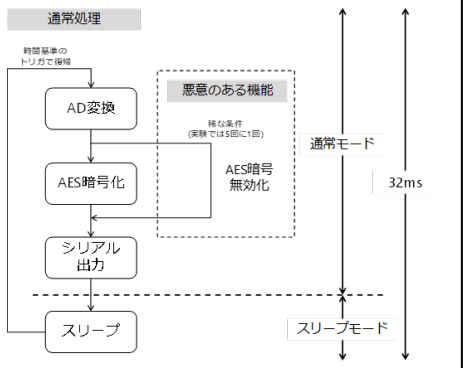


図 3-4-3：動作の概要

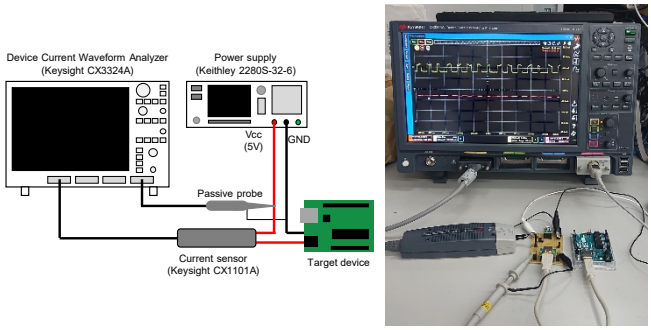
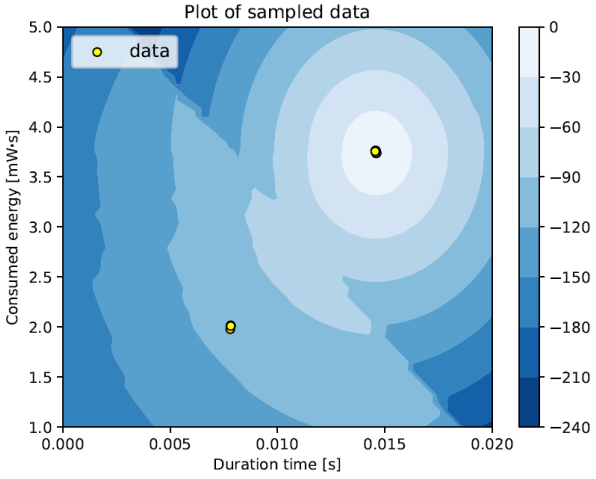


図 3-4-4：実験環境

図 3-4-4 に実験環境を示す。電源装置として Keithley 2280S-32-6 を使用し、この電源装置から対象となるマイクロコントローラに電源を供給する。実験では電源装置の出力電圧を 5V とし、最大電流を 0.4A と設定した。電流と電圧の測定には Keysight CX3324A を使用した。電圧はパッシブプローブを使用し、電流は電流センサ Keysight CX1101A を使用して測定した。オシロスコープで測定した消費電力にもとづき、図 3-4-2 のフローに従い、異常動作を検知した。

実験結果を図 3-4-5 に示す。図に示されるように、Arduino UNO を対象とした実験では LOF 値に明

確に異常な振る舞いを検出できる（セクション 5、 10、 15、 20、 25）。Arduino UNO は動作周波数が 16MHz となっており、比較的低速なため CMOS のスイッチングノイズも小さい。そのため、本実験評価でも悪意のある機能の動作を明確に検出することに成功した。



Active section	Duration [s]	Consumed power [mWs]	LOF
1	0.01461	3.74	-1.01
2	0.01457	3.76	-1.02
3	0.01457	3.76	-1.00
4	0.01459	3.74	-1.10
5	**0.00782**	**2.02**	**-109.57**
6	0.01457	3.77	-1.07
7	0.01458	3.75	-0.94
8	0.01461	3.74	-1.09
9	0.01458	3.75	-0.94
10	**0.00782**	**2.01**	**-109.67**
11	0.01459	3.75	-0.98
12	0.01461	3.74	-1.06
13	0.01458	3.75	-0.94
14	0.01457	3.76	-1.01
15	**0.00780**	**1.98**	**-110.45**
16	0.01460	3.74	-0.99
17	0.01457	3.76	-1.02
18	0.01457	3.76	-0.99
19	0.01460	3.74	-1.10
20	**0.00782**	**2.02**	**-109.57**
21	0.01456	3.76	-1.02
22	0.01459	3.75	-0.90
23	0.01460	3.74	-1.14
24	0.01459	3.76	-0.94
25	**0.00782**	**2.01**	**-109.69**
26	0.01459	3.75	-0.98
27	0.01460	3.74	-1.09
28	0.01458	3.76	-0.99
29	0.01455	3.76	-1.01

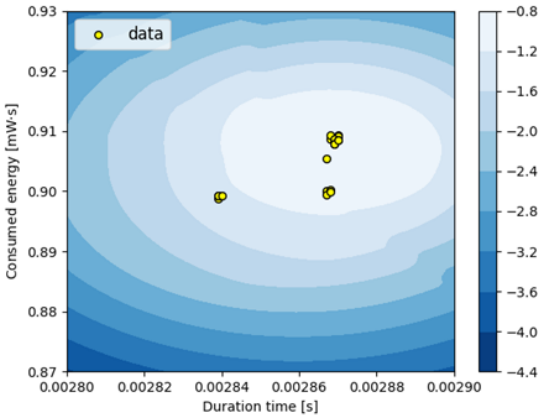
図 3-4-5：LOF による外れ値検知(1)

（4）Nucleo L476RG に関する異常動作検知

（3）と同じアプリケーションを別のマイクロコントローラでも動作させた。ここでは、組込みマイコンとして他の機能を多く持たず、マイクロコントローラとして使用可能な Nucleo L476RG（32 ビットマイクロコントローラ）を用いた。Nucleo L476RG でも暗号化機能をもったセンサロガーを想定し、通常モードでは、1) AD 変換、2) AES 暗号化、3) シリアル出力を行う。これらの処理を終えると、マイクロコントローラはスリープモードに移行する。実装したアプリケーションでは、50ms ごとに通常モードに復帰する。以上のアプリケーションに対し、悪意のある機能を挿入した。悪意のある機能では、5 回に 1 回 AES 暗号化処理を無効化し、AD 変換で得られた結果をそのままシリアル出力する。

実験結果を図 3-4-6 に示す。図に示されるように、Nucleo L476RG では、Arduino UNO と比較してアクティブ状態の継続時間が非常に短く、電力波形だけから視覚的に異常を検出するのは困難であるが、LOF 法を用いることで、異常動作の検出に成功した。図 3-4-6 では、異常な動作は左下のクラスタに当たり、他のクラスタよりも LOF 値が低いことが確認できる。本実験評価でも悪意のある機能の動作を明確に検出することに成功した。

これらの結果から、組込みプロセッサの動作モデル化と、電力波形に対する教師なし機械学習をベースとした異常動作検知によって、誤りなく異常検知を実現した。しかも複数種類の組込みプロセッサにおいて、異常検知が成功していることを実験的に示している。すなわち、課題Ⅱ－ア)において、令和元年度の目標を達成した。



Active section	Duration [s]	Consumed power [mWs]	LOF
1	0.002868	0.9087	-0.99
2	**0.002839**	**0.8987**	**-1.35**
3	0.002870	0.9094	-0.99
4	0.002869	0.9086	-0.99
5	0.002867	0.9000	-0.99
6	0.002868	0.9093	-0.99
7	**0.002839**	**0.8992**	**-1.34**
8	0.002869	0.9080	-0.99
9	0.002870	0.9092	-0.99
10	0.002869	0.9086	-0.99
11	0.002870	0.9085	-0.99
12	**0.002839**	**0.8992**	**-1.34**
13	0.002867	0.9054	-0.99
14	0.002867	0.8994	-0.99
15	0.002869	0.9079	-0.99
16	0.002869	0.9079	-0.99
17	**0.002840**	**0.8993**	**-1.31**
18	0.002868	0.9002	-0.99
19	0.002870	0.9086	-0.99
20	0.002868	0.8998	-0.99

図 3-4-6：LOF による外れ値検知(2)

３．５　課題Ⅱ-イ）AI/機械学習に基づく不正動作検知技術

イ）AI/機械学習に基づく不正動作検知技術
（１年目）単独の組込みマイコン等、比較的簡易な電子機器の動作のもと、見逃し確率を最小化するように、AI により外部情報から不正動作を検知する技術の基本検討を行う。

課題Ⅱ－イ）では、電子機器の状態を外部から観測するために電流を細かに計測・監視が可能となる事を課題とし、その実現のために、高精度アナログモジュールについて後述する精度を数値目標とし、目標精度の具体性の裏付け調査を実施した上で、回路の実装を試みる。

高精度アナログ計測モジュールの開発については、計測機器における電力プローブ及びアナログオペレータによる A/D 変換可能なレベルへの増幅が可能なハードウェアの検討を行い、モジュールを試作する。ハードウェアへの要求性能としては、既存の汎用計測器と遜色ない、±0.1mW の精度での電力の測定及び 1kHz 刻みでのサンプリング性能を求める。

並行して、不正動作検知モジュールの開発については、単独の組込みマイコン等、比較的簡易な電子機器の動作のもと、見逃し確率を最小化するように、外部情報から不正動作を検知する技術の検討を行う。

また、現状のハードウェアへの攻撃手法の調査を実施する。

(1) 高精度アナログ計測モジュール回路の実装

課題Ⅱ－アの研究において電子機器の外部から観測される電力データの特徴量を抽出するためは多機能・高性能の計測機器が利用されている。計測器に対して以下の機能を持つ基板を設計・実装する事により従来からの研究で用いられている計測器と比較し可搬性に優れ、性能的に遜色の無い点において高精度アナログ計測モジュールと考える事ができる。

- 従来同様の電力データ計測を可能とする
- 電力データ取得に特化した可用性の提供
- 計測器に比較して高い可搬性
- 最大で 1/20 程度のコスト、机上の省スペース化の実現

実図 3-5-1 に示すブロック図の回路を検討ならびに設計・実装を行った。この回路は電子機器の DC 電源に印加する電源に挿入される形となる。回路自体の電源は電子機器の DC 電源に影響のないように独立して別経路で用意する事となる。回路は DC 電源ならびに、そこに流れる電流を計測し、16bit の浮動小数点形式（BFP16）にて電力値を算出するものとなっている。

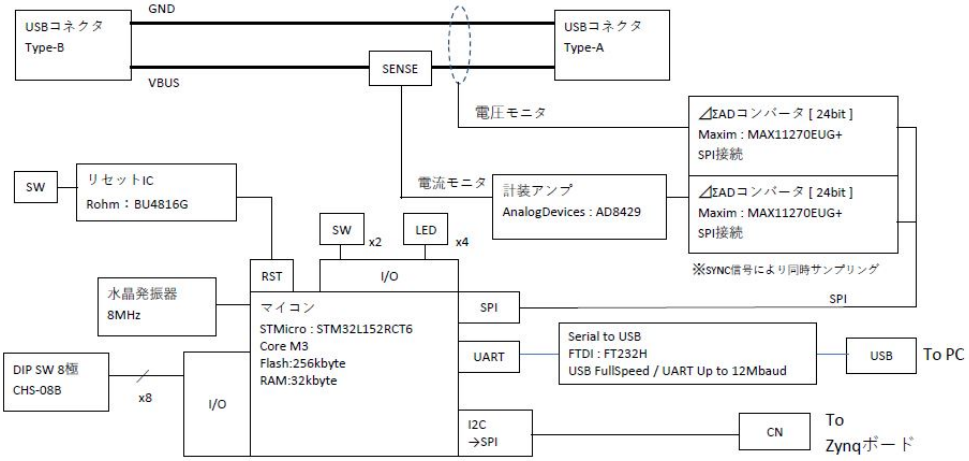


図 3-5-1 高精度アナログ計測モジュール回路ブロック図

(2) 高精度アナログ計測モジュールの評価

(1)で設計・実装した回路に対して Raspberry Pi を接続し、ネットワークに対して PING コマンドを実行しているときに、アウトカムとして定義したサンプリング精度、測定精度の目標値に達しているかの評価を実施した。まず、BFP16 の形式での PING の特徴量を捉えられるかの評価を実施した、そのときの BFP16 値から GNU plot を利用して作図した波形が図 3-5-2 となる。この時、サンプリング精度としては 5kHz で取り込む事に成功し、PING コマンド実行時の電力波形特徴量を捉える事が出来ているのが判る。

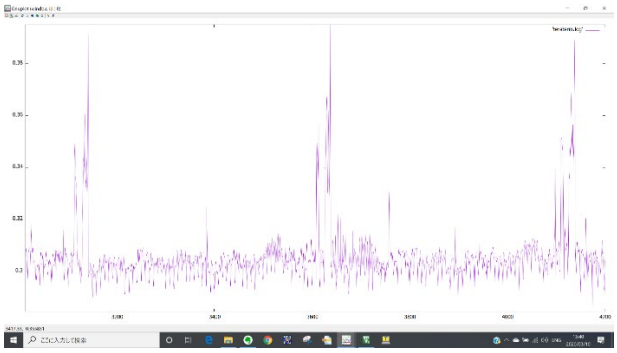


図 3-5-2 高精度アナログ計測モジュール PING 実行時の電力取り込み後の波形

次に BFP16 の測定精度について実際に小数点がどこまで丸められずに精度を出せているかの検算を以下のとおり実施した。PING コマンド実行時のピーク電力値が 0.339617W となっているとき、IEEE754 においては以下の形式の値となる。

IEEE754単精度(float, 32bit = 4Byte)では...

符号部	指数部 exponent 8							仮数部 fraction 23															
0	0	1	1	1	1	1	0	1	0	1	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	1	1	1
	3		E		A			D		E		2		4		7							

図 3-5-3 BFP16 値を IEEE754 形式で実数値→バイナリ変換した値

この結果を逆に浮動小数点値に検算すると以下のとおりとなる。

図 3-5-4 バイナリ値から実数値に変換したときの値

結果、令和元年度実施計画書に掲げた性能目標サンプリング精度 1kHz、測定精度±0.1mW に対してサンプリング精度 5kHz、測定精度±0.01mW という性能目標を上回る高精度アナログモジュールの実装を行う事ができた。

(3) 不正動作を検知する技術の検討

不正動作を検知する技術について、ニューラルネットワークを FPGA で実装する事を検討した。Xilinx 社製の FPGA ならびにその開発ツールを用いた場合、ニューラルネットワーク中で必要とされる積和演算器を FPGA で実装するために 1bit あたり FPGA の LUT を 1 個専有する形となる。課題Ⅱ－アに関する事前研究結果をベースに考えた場合、課題Ⅱ－アで用いられたニューラルネットワークの素子数が 300 個程度である。

結果、これを(1),(2)で検討～設計～実装した回路が出力する浮動小数点データ(16bit)で換算したとき、測定データ、重み付けデータ、出力データを持つ$3 \times 16 \times 300 = 14{,}400$となり、およそ 15,000 個の LUT によってニューラルネットワークを構成できるという検討結果になった。

(4) ハードウェアへの攻撃手法の調査

ハードウェアへの攻撃手法として不正なソフトウェアを電子機器に対し、ネットワークを経由して電子機器の脆弱性を狙うものが多かった。Virus Total を利用してマルウェアの種別を調査した結果として、1000 個近いマルウェアは 20 種類に分別された。分別されたマルウェアは、それぞれ Bot と呼ばれるもの、BackDoor/Dropper と呼ばれるものならびに破壊系とされるものであった。

電子機器のハードウェアに利用されるチップを模造するケースも存在している。中国の通販サイトで購入可能な IC チップに模造品が多いという情報から実際に中国の通販サイトと日本の正規代理店から、汎用通信 I/F チップを購入し、その違いについて調査を行った。今回の調査結果では、動作が不正になるという事はなく、図 3-5-5 にあるとおり見た目とチップ内部のパターンが異なるという点のみが判明した。

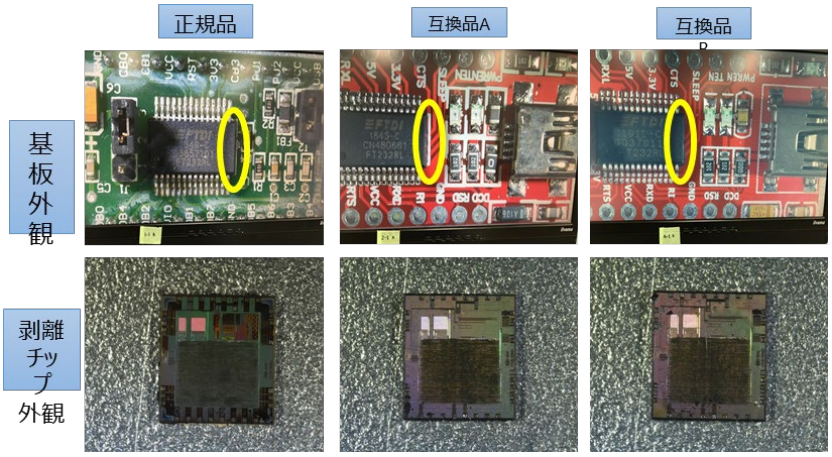


図 3-5-5 正規品と中国から購入した互換品

以上の結果から、不正に互換品を流通する事ができるという事実が不正回路の実在を裏付ける脅威として捉える事ができる。

4　政策目標（アウトカム目標）の達成に向けた取組みの実施状況

本研究開発ではアウトカム指標として以下を定める。

(1)　不正でない回路を不正と判定する誤検知率が 5%以下という条件のもと、不正回路を見逃す見逃し確率 10%以下

(2)　サプライチェーンでチップ脆弱性を検知、安全性を保証するしくみを構築

(3)　不正回路データベースの公開と「設計・製造におけるチップの脆弱性」検知の国際的ハブを形成

本研究開発は以上の取組みの成果を土台として、後述「5　政策目標（アウトカム目標）の達成に向けた計画」に述べるように、特許、認証スキームの活用などを目指す。

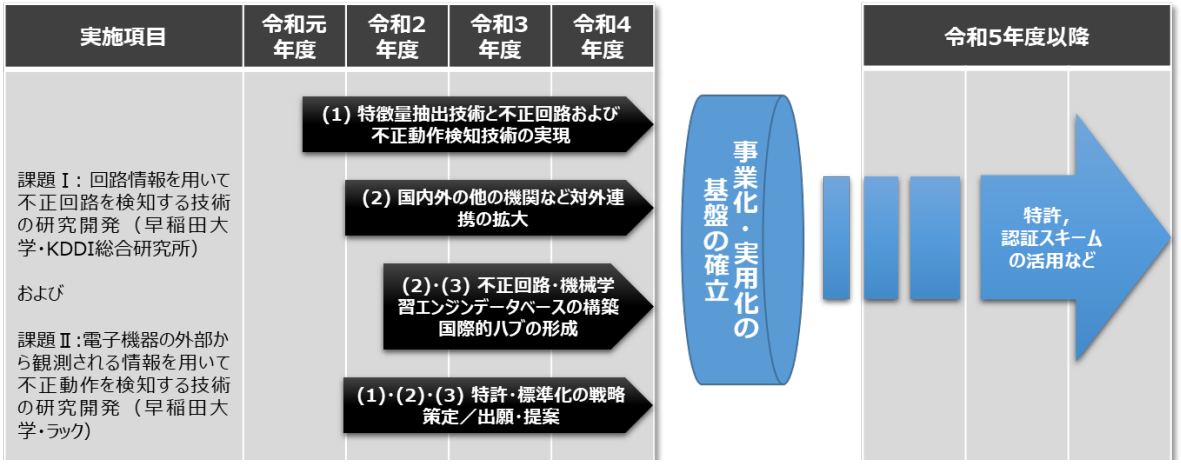


図 3-4 本研究開発のアウトカム指標

令和元年度における具体的な進捗については次の通りである。上記 1 について、本研究開発における技術が社会実装として有用なものとなるか否かはその数値目標の設定と達成にかかっている。この点、令和元年度、誤検知率の妥当性の評価を行った。令和元年度に達成した数値目標（TPR80%以上、TNR95%以上）であっても、ハードウェアトロイ特定に実用レベルで利用可能性がある（ハードウェア脆弱性の検証を実運用

している半導体設計メーカの意見による)。もっとも、実際は、個別の結果を見て判断する必要があることを踏まえ、今後、産業界との実利用を想定し議論を継続することを前提に、当面、「TNR を 95%以上、TPR を 90%以上」を目標とすることが妥当だ、との方向性を示した。この方向性は本研究開発における有識者の研究開発運営委員会の賛同を得た。また、前述の通り、不正でない回路を不正と判定する誤検知率を 5%以下にした上で、不正回路を見逃す見逃し確率を 15%強程度にすることが達成されており、令和元年度の目標が達成されている。

上記 2 について、ハードウェアトロイ検知の出口戦略の調査も行った。たとえば、回路情報を用いた不正回路検知の実例として、東芝情報システムによるものがある(https://www.tjsys.co.jp/info/news/003554.htm を参照)。本研究開発における技術は、こうした既存のツールに対して、機械学習等 AI 技術を用いることで、より高度化したものとなり、半導体チップのサプライチェーンのセキュア化に貢献する。その上で、今後、上記のような認証スキームに活用されるモデルを出口戦略として、具体的に検討を進める。

上記 3 について、ISO/IEC JTC 1/SC 27/WG3 への寄書のドラフトを作成した。これは、中国が提案している「Hardware Security Monitoring Framework」(Study Period)のスコープを拡張し、回路情報に基づく不正回路の検知（本研究開発における課題 I）も含めることを提案するものである。そして、ISO/IEC SC27 会合（2020 年 4 月）にて、本プロジェクトの成果を紹介した。また、同 WG3 で検討されているハードウェアセキュリティに関する SP の共同ラポーターに就任した。これにより、本研究開発の成果を主導的に標準化提案に盛り込むことが可能となった。

5　政策目標（アウトカム目標）の達成に向けた計画

本施策のアウトカム目標として、大きく以下の 3 点を見込んでいる。

1. 設計・製造におけるチップの脆弱性検知手法の確立
2. サプライチェーン上での運用技術の確立と社会実装の加速
3. 特許取得・業界標準化等を通じた我が国の国際競争力強化

加えて、本研究開発においては、次の出口戦略も見込んでいる。

4. 1.～3.の成果に基づく長期的目標として、サプライチェーンでチップ脆弱性を検知、安全性を保証するしくみの実用化・事業化

具体的には、本施策の成果を活用することで、標準化・ガイドライン化や認証スキームを確立し、セキュアなサプライチェーンを構築する。

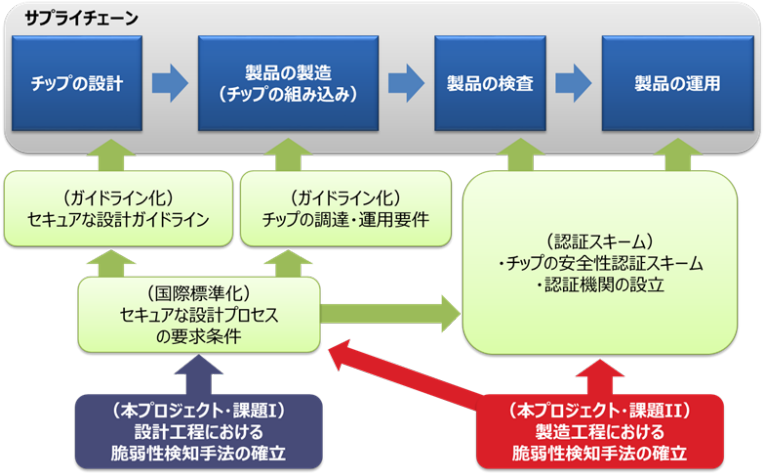


図 3-5 本研究開発のアウトカム目標

以上の実用化・事業化の具体策は、研究開発運営委員会、プログラムディレクタ、及びビジネスプロデューサの助言・統括の下、順次、検討・具体化する。

以上の事業化・実用化により、我が国の半導体市場に有利な影響を与えることが予想される。すなわち、現在、世界の半導体の市場規模は 5000 億ドル弱と言われている。一方、世界全体における日本の半導体シェアは 7%程度となる。世界の半導体の市場規模は、今後も 10%程度の伸びを示すと考えられ、本研究開発によって「チップの信頼性」が高まることによって、アジア太平洋地域の半導体シェアに食い込み、例えば我が国の半導体シェアが 2000 年代前半の 20%となれば, 年間 5000 億ドル×(20%-7%)=650 億ドルの効果が見込まれる。加えて、世界的には半導体市場は拡大しており、それに応じてその効果も高まるものと予想される。

６　査読付き誌上発表論文リスト

なし

７　査読付き口頭発表論文（印刷物を含む）リスト

なし

８　その他の誌上発表リスト

なし

９　口頭発表リスト

[1] 井上智貴, 長谷川健人, 戸川望, "ニューラルネットワークを用いたハードウェアトロイ検出における局所性の応用に関する一考察," 電子情報通信学会総合大会, A-20, 2020 年 3 月 18 日（広島県広島市）.

[2] 井上智貴, 長谷川健人, 戸川望, "トリガ回路の性質にもとづく特徴量を利用したニューラルネットワークによるハードウェアトロイ識別," 電子情報通信学会技術報告 VLD2019-133, HWS2019-106, 2020 年 3 月 6 日（沖縄県那覇市）.

[3] 野澤康平, 長谷川健人, 披田野清良, 清本晋作, 橋本和夫, 戸川望, "ニューラルネットワークを用いたハードウェアトロイ識別に対する敵対的サンプル攻撃の順序回路への応用," 電子情報通信学会・2020 年暗号と情報セキュリティシンポジウム, 2020 年 1 月 30 日（高知県高知市）.

[4] 栗原樹, 長谷川健人, 戸川望, "ゲートレベル IP コアを対象とした機械学習によるハードウェアトロイ識別の評価," 電子情報通信学会・2020 年暗号と情報セキュリティシンポジウム, 2020 年 1 月 29 日（高知県高知市）.

[5] 高崎和成, 長谷川健人, 木田良一, 戸川望, "シングルボードコンピュータを対象とした電力解析にもとづくアプリケーション電力の抽出," 電子情報通信学会・2020 年暗号と情報セキュリティシンポジウム, 2020 年 1 月 29 日（高知県高知市）.

[6] 長谷川健人, 木田良一, 戸川望, "シングルボードコンピュータを対象とした電力振る舞いの解析と異常動作検知への応用," 電子情報通信学会･2020 年暗号と情報セキュリティシンポジウム, 2020 年 1 月 29 日（高知県高知市）.

[7] 長谷川健人, 木田良一, 戸川望, "電力解析にもとづく異常動作検知装置の実装と評価," 電子情報通信学会・2020 年暗号と情報セキュリティシンポジウム, 2020 年 1 月 29 日（高知県高知市）.

１０　出願特許リスト

なし

１１　取得特許リスト

なし

１２　国際標準提案・獲得リスト

なし

１３　参加国際標準会議リスト

なし

１４　受賞リスト

なし

１５　報道発表リスト

（１）報道発表実績

なし

（２）報道掲載実績

なし

１６　ホームページによる情報提供

[1] “チップ脆弱性検知手法の研究開発採択”、2019 年 10 月 25 日、早稲田大学、https://www.waseda.jp/top/news/66858

[2] “ハードウェアチップの脆弱性検知手法の研究開発に採択”、KDDI 総合研究所、2019 年 10 月 25 日 https://www.kddi-research.jp/newsrelease/2019/102501.html

[3] “ハードウェアチップの脆弱性検知手法の研究開発に採択”、株式会社ラック、2019 年 10 月 25 日 https://www.lac.co.jp/news/2019/10/25_press_01.html

研究開発による成果数

	令和元年度	合計
査読付き誌上発表論文数	0件（ 0件）	0件（ 0件）
査読付き口頭発表論文数（印刷物を含む）	0件（ 0件）	0件（ 0件）
その他の誌上発表数	0件（ 0件）	0件（ 0件）
口頭発表数	7件（ 0件）	7件（ 0件）
特許出願数	0件（ 0件）	0件（ 0件）
特許取得数	0件（ 0件）	0件（ 0件）
国際標準提案数	0件（ 0件）	0件（ 0件）
国際標準獲得数	0件（ 0件）	0件（ 0件）
受賞数	0件（ 0件）	0件（ 0件）
報道発表数	0件（ 0件）	0件（ 0件）
報道掲載数	0件（ 0件）	0件（ 0件）